

miRTarVis: microRNA 와 mRNA 표현형 프로파일 시각화 분석 도구에 관한 연구

정대경⁰¹ 서진욱¹

¹서울대학교 컴퓨터공학부

rati@hcil.snu.ac.kr, jseo@snu.ac.kr

miRTarVis: a standalone tool for visual analysis of microRNA and mRNA expression profile

Daekyoung Jung⁰¹, Jinwook Seo¹

¹School of Computer Science and Engineering

요 약

본 논문에서는 microRNA (miRNA)-mRNA 의 표현형 프로파일 데이터의 탐색 및 분석을 돕기 위한 인터랙티브 시각화 분석 도구인 miRTarVis를 제시한다. 기존의 miRNA 표적 예측 알고리즘은 과도한 표적 후보를 제시하고, miRNA-mRNA 표현형 프로파일 데이터와 결합하기 어려웠다. 몇몇 웹 기반의 분석 도구들이 miRNA-mRNA 이 문제를 해결하기 위해 제시되었지만, 효율적인 해결책이 되지 않았다. miRTarVis를 통해 사용자는 miRNA-mRNA 표현형 데이터를 효과적으로 질의, 탐색하고, miRNA 표적 예측 알고리즘 및 연관성 분석방법과 결합한 인터랙티브한 시각적 분석을 수행하여 이 문제를 해결할 수 있다. 결과적으로, miRTarVis 에서 제공하는 작업 흐름에 따라 miRNA-mRNA 표현형 데이터를 분석하면 miRNA 의 표적에 대한 가설 설정에 도움을 받을 수 있을 것이다.

1. 서 론

microRNA (miRNA) 는 mRNA 의 분해를 촉진하고 번역을 억제하여 후전사조절을 수행하는 짧은 (~22nt) RNA 이다. [1] miRNA 의 표적 mRNA 을 실험적으로 검증하는 것이 어렵기 때문에, miRNA 의 염기 서열 정보와 보존성 정보를 활용한 예측 알고리즘들이 개발되었다.

그러나 이들 예측 알고리즘은 민감도 (sensitivity) 가 낮아서, [2] 이를 해결하기 위한 해결책이 필요하다. 이 문제를 해결하기 위한 한 방법은 알고리즘에서 제시하는 예측의 정확성을 나타내는 예측 점수에 따라 높은 점수를 가지는 miRNA-mRNA 상호작용만 걸러내는 방법이지만, 대부분의 알고리즘에서 이 기능을 명시적으로 제공하지 않는다. 또 다른 방법은 여러 개의 알고리즘에서 겹치는 miRNA-mRNA 상호작용을 찾아내는 것이지만, 독립적인 예측 알고리즘을 결합하는 도구가 많지 않다. miRNA-mRNA 표현형 데이터와 예측 결과를 결합하는 것도 한 방법이지만, 연구자 입장에서 이를 쉽게 수행할 수 있도록 도와주는 도구가 부족하였다.

이 문제들을 해결하기 위해 몇몇 웹 기반의 도구들 ([3]) 이 다양한 예측 알고리즘, 실험 데이터와 예측 알고리즘의

결합과 인터랙티브한 시각화를 지원한다. 그러나 이들은 사용 가능한 예측 알고리즘의 숫자의 제약을 가지고 있거나, 유의미한 miRNA 와 mRNA 의 표현형 값을 실시간으로 인터랙티브하게 필터링하는 등의 기능을 가지고 않다. 또한 miRNA-mRNA 표현형 데이터에 대한 연관성 분석과 miRNA 표적 예측 알고리즘을 결합시켜 예측의 신뢰성을 높이는 기능이 부족하다. 이에 따라 우리는 miRNA-mRNA 표현형 프로파일을 miRNA 예측 알고리즘 (TargetScanHuman [4], microRNA.org [5], miRDB [6], PITA [7], Pictar [8]) 및 실험적으로 검증된 miRNA 표적 데이터베이스 (miRecords [9], miRTarBase [10]) 와 결합하고, 표현형 프로파일에 대한 연관성 분석을 통해 시각적 분석을 돕는 독립형 도구 miRTarVis를 개발하였다.

2. 구현 및 도구 명세 기술

먼저 miRNA-mRNA 표현형 프로파일 데이터 시각화 분석 도구의 목표는 서로 다른 조건에서 차별적으로 표현된 (differentially expressed) miRNA를 검출하고 이들의 가능한 표적 후보들을 평가하는 것이라고 가정하고, 이에 따라 miRTarVis 의 사용시 작업흐름을 설계하였다. (그림 1. B)

이 논문은 0000년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. NRF-2014R1A2A2A03006998).

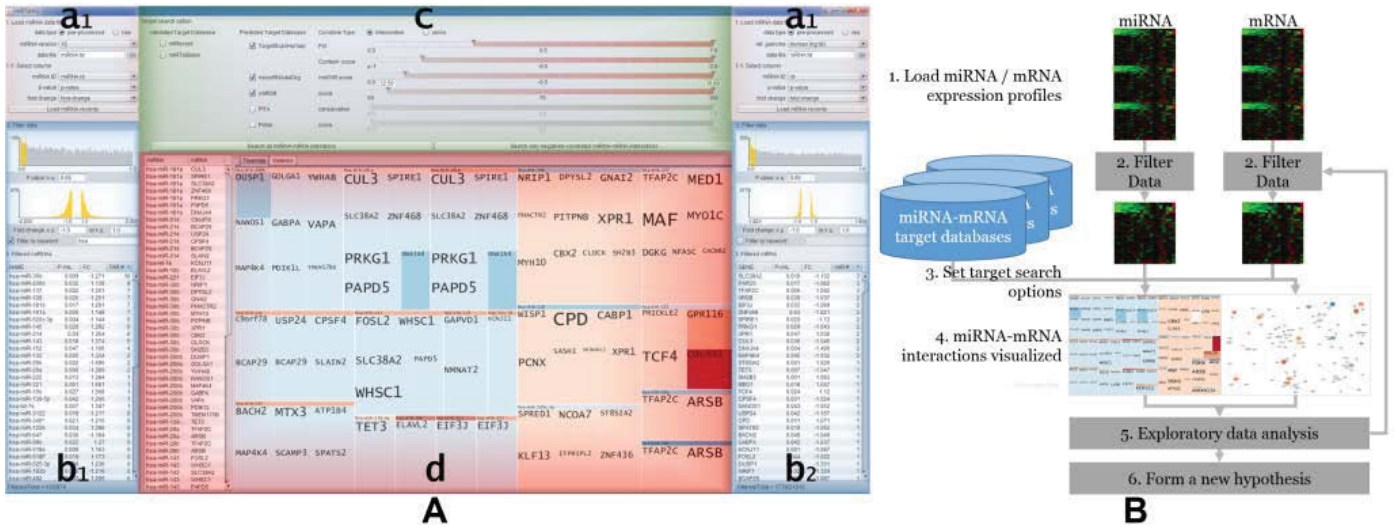


그림 1. A: miRTarVis, a₁, a₂: 데이터 로드, b₁, b₂: 데이터 필터, c: 표적 예측 설정, d: 시각화 B: 작업흐름

miRTarVis의 특징은 다음과 같다. miRTarVis는 탭으로 구분된 텍스트 파일을 입력으로 받는다. 입력은 식별자(miRNA-mRNA ID)와, 사전에 계산된 p값과 변화량(fold change)으로 이루어진 데이터이거나, 실험 전후 여러 개의 샘플에서 측정된 표현형 값으로 이루어진 데이터이다. 이 경우 p값과 변화량은 자동으로 계산된다. 데이터를 로드하면 p값과 변화량의 분포를 보여주는 히스토그램이 나타나고, miRNA와 mRNA의 p값과 변화량의 값들이 히스토그램 아래의 테이블에 나타난다. miRNA와 mRNA에 대한 더 자세한 정보를 제공하기 위해 외부 웹 서비스 링크들(miRBase [11], TargetScanHuman, miRDB, miR2Disease [12], UCSC Genome browser, NCBI, Gene card)을 제공한다.

예측 알고리즘은 선별적으로 선택된다. 사전에 편집된 데이터베이스가 miRTarVis 내에 삽입되어 있다. 사용자가 여러 개의 표적 예측 알고리즘을 선택한 경우에는 모든 선택한 예측 알고리즘을 만족시키는 miRNA-mRNA 상호작용들만 선택된다. 사용자는 표적 예측에 사용되는 예측 점수를 선택할 수 있다. 이 점수의 크기와 방향은 예측 알고리즘마다 크기와 방향이 다르지만 공통적으로 필터 인터페이스 상에서 더 신뢰도가 높은 점수 범위가 오른쪽에 위치하도록 설정되었다. 일반적으로 miRNA는 표적 mRNA의 발현을 억제한다고 알려져 있기 때문에, miRNA-mRNA 상호작용 중 같은 방향의 경향성을 보이는 상호작용들은 걸러낼 수 있는 옵션도 지정할 수 있다.

miRTarVis는 이렇게 걸러진 miRNA-mRNA 표현형 프로파일과 선택된 표적 예측 검색 설정에 따라서 표현형 프로파일과 표적 예측 결과를 통합시키고, 결과를 표, 노드-링크 네트워크 다이어그램, 트리맵 시각화로 보여준다. (그림 1. A) 변화량을 노드의 색상으로 보여주기 때문에, 트리맵 시각화는 miRNA-mRNA 관계를 정확히 보여줄 수 있다. 특히 결과로 표현될 miRNA-mRNA 상호관계의 숫자가 많은 경우 트리맵 시각화가 더욱 유용할 것으로 기대된다. 트리맵

시각화뿐만 아니라, 전통적인 네트워크 다이어그램과 표 또한 miRNA-mRNA 상호작용들을 나타내는데 사용된다. 이들은 아주 인터랙티브하다. 사용자가 원하는 miRNA 또는 mRNA를 선택하면 상세 정보가 나타나게 된다. 만약 사용자가 miRNA 또는 mRNA를 필터링 하기 위한 cutoff 값이나 표적 예측의 설정 값을 바꾸면 시각화 결과는 그에 따라 직접 업데이트된다. 또한 원하는 영역을 줌 인/아웃 하거나 더 자세한 외부 데이터베이스를 참조하는 것도 가능하다. 이를 통해 사용자의 데이터를 탐색을 돕는다.

3. 결론

이 논문에서는 miRNA-mRNA 표현형 프로파일을 miRNA 표적 예측 알고리즘과 결합하여 분석하고 인터랙티브한 시각화를 제공하는 도구인 miRTarVis를 소개하였다. miRTarVis는 다른 도구들과 비교하여 miRNA 연구에 다음과 같은 장점을 가지고 있다. miRTarVis는 여러 개의 표적 예측 알고리즘을 동시에 사용할 수 있으며, 표적 예측 점수로 miRNA-mRNA 상호작용을 필터링 할 수 있고, 최초로 트리맵 시각화를 miRNA-mRNA 상호작용 표현에 사용했으며, 다른 도구보다 향상된 상호작용을 제공한다. 따라서 miRTarVis는 연구자가 miRNA-mRNA 표현형 프로파일로부터 가설을 설정하는 데 도움을 줄 것으로 예측된다.

참고문헌

[1] Bartel, D. P. MicroRNAs: Genomics, biogenesis, mechanism, and function. *Cell*, 116(2), 281 – 297. 2004
 [2] Sethupathy, P., Megraw, M., and Hatzigeorgiou, A. G. A guide through present computational approaches for the identification of mammalian microRNA targets. *Nature Methods*, 3(11), 881 – 886. 2006
 [3] Nam, S., Li, M., Choi, K., Balch, C., Kim, S., and Nephew, K. P. MicroRNA and mRNA integrated analysis (mmia): a web tool for

examining biological functions of microrna expression. *Nucleic Acids Research*, 37(suppl 2), W356–W362. 2009

[4] Lewis, B. P., Burge, C. B., and Bartel, D. P. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microrna targets. *Cell*, 120(1), 15 – 20. 2005

[5] Betel, D., Wilson, M., Gabow, A., Marks, D. S., and Sander, C. The microrna.org resource: targets and expression. *Nucleic Acids Research*, 36(suppl 1), D149–D153. 2008

[6] Wang, X. mirdb: A microrna target prediction and functional annotation database with a wiki interface. *RNA*, 14(6), 1012–1017. 2008

[7] Kertesz, M., Iovino, N., Unnerstall, U., Gaul, U., and Segal, E. The role of site accessibility in microrna target recognition. *Nat Genet*, 39(10), 1278–1284. 2007

[8] Anders, G., Mackowiak, S. D., Jens, M., Maaskola, J., Kuntzack, A., Rajewsky, N., Landthaler, M., and Dieterich, C. dorina: a database of rna interactions in post-transcriptional regulation. *Nucleic Acids Research*, 40(D1), D180–D186. 2012

[9] Xiao, F., Zuo, Z., Cai, G., Kang, S., Gao, X., and Li, T. mirecords: an integrated resource for micrornatarget interactions. *Nucleic Acids Research*, 37(suppl 1), D105–D110. 2009

[10] Hsu, S.-D., Lin, F.-M., Wu, W.-Y., Liang, C., Huang, W.-C., Chan, W.-L., Tsai, W.-T., Chen, G.-Z., Lee, C.-J., Chiu, C.-M., Chien, C.-H., Wu, M.-C., Huang, C.-Y., Tsou, A.-P., and Huang, H.-D. mirtarbase: a database curates experimentally validated microrna target interactions. *Nucleic Acids Research*, 39(suppl 1), D163–D169. 2011

[11] Kozomara, A. and Griffiths-Jones, S. mirbase: integrating microrna annotation and deep-sequencing data. *Nucleic Acids Research*, 39(suppl 1), D152–D157. 2011

[12] Jiang, Q., Wang, Y., Hao, Y., Juan, L., Teng, M., Zhang, X., Li, M., Wang, G., and Liu, Y. mir2disease: a manually curated database for microrna deregulation in human disease. *Nucleic Acids Research*, 37(suppl 1), D98–D104. 2009